residues comprising a particular protein, can then be manipulated to produce other models, such as those depicting alpha-carbon atom representations, all heavy atom representations, and even all atom representations.

In alternative embodiments, such representations can be used to determine protein function using one or more three-dimensional templates correlated with particular biological functions, and they can also be used to identify functionally important regions in a protein. *See, e.g.*, Kasuya, A. and Thornton, J.M., *J. Mol. Biol.*, vol. 286: 1673-1691 (1999); Wallace, *et al.* (*Protein Science*, vol. 5:1001-1013 (1996); Bone, *et al.*, *Biochemistry*, vol. 30:10388-10398 (1991); Barth, *et al.* (1993) *Drug Design and Discovery*, vol. 10:297-317; Gregory, *et al.* (1993), *Protein Eng.*, vol. 6, no. 1:29-35; Artymiuk, *et al.* (1994), *J. Mol. Biol.*, vol. 243:327-344; and Fischer, *et al.* (1994), *Protein Sci.*, vol. 3:769-778). A particularly preferred approach employs functional site descriptors (FSDs). *See* U.S.S.N. 09/322,067, filed May 27, 1999. Using FSDs, prediction of a protein's biological function requires only an approximation of the three-dimensional orientation of two or more amino acid residues in a region responsible for the particular function of the protein under investigation. Broadly, FSDs define spatial configurations for protein functional sites that correspond with particular biological functions, and it is known that function derives from structure. FSDs provide three-dimensional representations of protein functional sites, for example, ligand binding domains (*e.g.*, domain that bind a ligand, for example, a substrate, a co-factor, or an antigen), protein-protein interaction sites or domains, and enzymatic active sites.

A functional site descriptor typically comprises a set of geometric constraints for one or more atoms in each of two or more amino acid residues comprising a functional site of a protein. Preferably, the atoms are selected from the group consisting of amide nitrogens, α-carbons, carbonyl carbons, and carbonyl oxygens within a polypeptide backbone, β-carbons of amino acid residues, and pseudoatoms, *e.g.*, a side chain center of mass.

The geometric constraints of an FSD preferably are selected from the group consisting of an atomic position specified by a set of three dimensional coordinates, an interatomic distance (or range of interatomic distances), and an interatomic bond angle (or range of interatomic bond angles). When a geometric constraint refers to atomic position, reference is typically made to a set of three-dimensional coordinates. Such constraints can relate to RMSDs. Other geometric constraints concern interatomic distances, preferably interatomic distance ranges, or interatomic bond angles, preferably interatomic bond angle ranges.

In some embodiments, an FSD can also include one or more conformational constraints that refer to the presence of a particular secondary structure, for example, a helix, or location, for example, near the amino or carboxy terminus of a protein. FSDs can be implemented in electronic form, so that they can be used in computerized methods. Typically, functional site descriptors comprising two to about 50 or more geometric constraints can be developed for a particular biological function. In many embodiments, the number of geometric constraints in an FSD is from about 4-25, often from about 5-20.

As indicated above, FSDs can be built for any type of protein function. Functions of particular interest include enzymatic activities. At present, more than 180 different enzymatic activities have been classified, and are listed by enzyme name in the following table. The particular classification of an enzyme listed in the following table is defined in accordance with the enzyme classification system as described in, *e.g.*, *Enzyme Nomenclature*, NC-IUBMB, Academic Press, New York, New York (1992), and at www.biochem.ucl.ac.uk/bsm/enzymes/index.html.

E.C. Number	Enzyme Name
1.1.1.2	Alcohol dehydrogenase (NADP+)
1.1.1.21	Aldehyde reductase
1.1.1.27	L-lactate dehydrogenase
1.1.1.28	D-lactate dehydrogenase

E.C. Number	Enzyme Name
1.1.1.29	Glycerate dehydrogenase
1.1.1.34	HMG-CoA reductase
1.1.1.42	Isocritrate dehydrogenase (NADP+)
1.1.1.49	Glucose-6-phosphate 1-dehydrogenase
1.1.1.50	3-alpha-hydroxysteroid dehydrogenase (B-specific)
1.1.1.53	3-alpha(or 20-beta)-hydroxysteroid dehydrogenase
1.1.1.62	Estradiol 17 beta-dehydrogenase
1.1.1.95	Phosphoglycerate dehydrogenase
1.1.1.159	7-alpha-hydroxysteroid dehydrogenase
1.1.1.184	Carbonyl reductase (NADPH)
1.1.1.206	Tropine dehydrogenase
1.1.1.236	Tropinone reductase
1.1.1.252	Tetrahydroxynaphthalene reductase
1.1.3.7	Aryl-alcohol oxidase
1.1.3.15	(S)-2-hydroxy-acid oxidase
1.1.99.8	Alcohol dehydrogenase (acceptor)
1.2.1.2	Formate dehydrogenase
1.2.1.5	Aldehyde dehydrogenase (NAD(P)+)
1.2.1.8	Betaine-aldehyde dehydrogenase
1.2.1.12	Glyceraldehyde 3-phosphate dehydrogegnase (phosphorylating)
1.2.3.3	Pyruvate oxidase
1.3.99.2	Butyryl-CoA dehydrogenase
1.4.1.2	Glutamate deydrogeanse
1.4.1.3	Glutamate dehydrogenase (NAD(P)+)
1.4.3.3	D-amino acid oxidase
1.4.3.6	Amine oxidase (copper-containing)
1.5.1.3	Dihydrofolate reductase
1.6.4.2	Glutathione reductase (NADPH)